Small Molecules in Biomedical Research at the University of Oxford

April 2013

Paul Brennan,[1,2*] Hannah Lingard,[2] Rachel MacCoss,[3] Brian Marsden[2]

Contents

1. *Executive Summary*
2. *Introduction*
3. *Benchmarking and Target Discovery*
   - *3.1 Chemogenomic Sets*
   - *3.2 Quality Control of Compound Sets*
4. *Lead Discovery and Phenotypic Screening*
5. *Lead Optimization*
6. *Data Capture and Sharing*
7. *Metabolomics*
8. *Strategic Recommendations*
   - *8.1 Screening Set QC*
   - *8.2 Compound and Screening Database*
   - *8.3 Outsourcing Lead Optimisation Assays*
   - *8.4 Metabolomics Capabilities*
9. *Appendix*

[1] Target Discovery Institute, Nuffield Department of Medicine

[2] Structural Genomics Consortium, Nuffield Department of Medicine

[3] Department of Chemistry

*To whom correspondence should be addressed:
paul.brennan@sgc.ox.ac.uk, telephone: 01865 617 574.

1. Executive Summary

Synthetic small, drug-like molecules are used throughout the University of Oxford for biomedical research, both as tools to disentangle disease relevant biological pathways and at all stages in the drug discovery pathway from lead to approved medicine. Married with the use of synthetic compounds in biomedical research is metabolomics: the quantitative analysis of endogenous small molecules on a large scale in order to measure the state of cellular or whole organism physiology or pathology. For small molecules, the needs of researchers within the University vary greatly based on the number of compounds used (single compounds to tens of thousands) and the systems studied. In metabolomics research, the breadth is great with the major differences due to the source of samples (cellular extracts to clinical isolates), method of analysis (MS versus NMR) and pathways studied. In order to understand current practices and to identify gaps in the University's capabilities, we interviewed 28 researchers across 22 departments in the University's MPLS and MS divisions from computer scientists to practicing clinicians (Appendix, Table A1). Although the range of responses was vast from "we don't do any of these things" to hour-long discussions, four common areas for improvement in how the University uses small molecules in biomedical research emerged:

1. *Quality control of large physical collections of small molecules.*
2. *Easier University-wide access to compound collections, assay data and assay expertise.*
3. *Better understanding of how small molecules testing in lead optimization can be outsourced.*
4. *Increased capacity and expertise for metabolomics studies in all areas.*

2. Introduction

Small molecules are used extensively within the University in biomedical research on a scale ranging from a single molecule as a tool to interrogate cellular pathways, to screening of thousands of compounds in phenotypic cellular assays or animal studies to discover a unique combination of compound structure and desired effect (Figure 1).

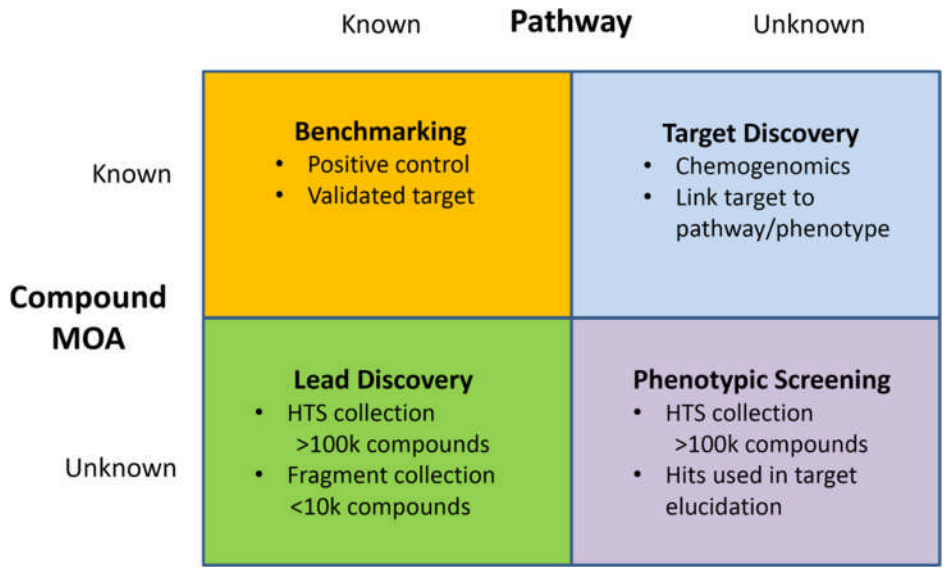

Figure 1. Role of small molecules in biomedical research

In the simplest case, a compound with a known mechanism of action (MOA) is tested in a known system to verify a precedented effect. In this case, **Benchmarking** (Figure 1), the compound is used as a standard or positive control to verify the system. Benchmarking is important as it provides confidence that the assay conditions or animal model are appropriate to use in discovery mode to find novel molecules, biological targets or disease pathways.

When moving from **Benchmarking** to **Target Discovery**, the number of small molecules used increases from a handful to thousands of compounds. The use of well characterised small molecules to interrogate biological function for target discovery is known as chemical genetics. A chemogenomics set of compounds is collected where each compound ideally has a well-established pharmacological mode of action on a small set of biological targets. The compounds are screened for a phenotypic effect and the compound is used to link the target to the effect. As compounds are rarely selective for a single target and may have unknown pharmacologies, the ideal chemogenomics set will be populated with chemically distinct compounds with reported identical pharmacology. An effect seen with multiple compounds sharing the same pharmacology links the target to the phenotypic effect with confidence.

Moving from **Benchmarking** down an orthogonal axis to **Lead Discovery** requires even larger compound collections. In this discovery mode, a target has already been identified through other means (e.g. RNAi, GWAS) and a small molecule hit is required to implement a chemical probe or drug discovery project. As the number of possible small molecules is vast, a high-throughput screen (HTS) of at least 100k molecules is needed to have any chance of finding a hit. Alternatively, a fragment screen of hundreds or thousands of much smaller molecules gives the same chance of hit discovery but is limited to biophysical assays for the target.

In the final mode, neither the target or molecule conferring a desired phenotype is known, but induction of phenotype is amenable to high-throughput screening, so both can be discovered from a **Phenotypic** assay. A large collection of small molecules (10k – 100k) is screened in the phenotypic assay and resultant hits are then used to identify the biological target responsible for the desired phenotype.

At first thought, metabolomics in drug discovery seems distinct from the cases outlined so far, but in fact metabolic studies share synergies with small molecules in target discovery, hit discovery and phenotypic screening. Metabolomic analysis requires a collection of small molecule standards and sensitive analytic techniques to quantify effects on a cell's or organism's metabolome.

The scope of this audit was to determine the availability and uses of small molecules and metabolomics capabilities in biomedical research within the University of Oxford as outlined in Figure 1. Related topics which lie out of the scope are design, synthesis and purification of small molecules, formulation research, biologics (proteins, antibodies, peptides and hybrids), RNAi, gene therapy and vaccines.

3. Benchmarking and Target Discovery

3.1 Chemogenomic Sets

Most of the small molecules used within the University are either used for benchmarking or target discovery. A majority of research groups surveyed use a small number of compounds in cellular systems to see a chemical genetic effect in disease models. More comprehensive collections are usually focused on approved drugs. This provides an advantage in terms of ease of development of a molecule for a newly discovered target or indication as the *in vivo* profile and toxicity of these agents is usually well established. The disadvantage of using only known drugs is that they are limited in the number of targets they interrogate. Table 1 shows the chemogenomics compound sets used by researchers surveyed in this audit. Although the Pharmakon, Prestwick, Johns Hopkins and BML-2842 collections focus on approved drugs, the other sets have additional known bioactive compounds. The Published Kinase Inhibitor Set (PKIS) contains a large number of kinase inhibitors

discovered by GSK and the epigenetic set contains a small number of specific compounds targeting a range of epigenetic proteins. These two sets are of particular interest as they move beyond approved drugs and incorporate compounds with novel pharmacology that have not necessarily been used clinically. Most sets have little overlap so all should be considered when screening in target discovery (see Figure A1).

Name	Number of compounds	Source	Owner	Description
Pharmakon	1600	MSDI	TDI	1600 known drugs from US and International sources
Spectrum Collection	1600	MSDI	TDI	Approved drugs (60%), natural products (25%), other bioactives (15%)
Clinical Compound Library	1500	Johns Hopkins	TDI	FDA and foreign-approved drugs
LOPAC	1280	Sigma	TDI	Library of Pharmacologically Active Compounds
Prestwick Chemical Library	1200	Prestwick	TDI	FDA approved drugs
Clinical Collections	731	NIH	TDI	Drugs tested in Phase I – III
BML-2842	640	Enzo	TDI	FDA approved drugs
BML-2840	480	Enzo	TDI	ICCB known bioactives library
PKIS	367	GSK	TDI	Published Kinase Inhibitor Set
DTP Approved Oncology Drugs	89	NCI/NIH	TDI	Most of the current FDA-approved anticancer drugs
Epigenetics Set	40	Custom	Botnar, SGC	Hand-curated set of epigenetic inhibitors

Table 1. Chemogenomics Sets

3.2 Quality Control of Compound Sets

Quality control is an important part of using chemogenomic compound sets in target discovery as small molecules have a limited shelf-life due to precipitation and decomposition when

stored as screening solutions in solvent. For groups that use only a small number of purchasable compounds, QC is not a major issue. As one PI said, "When it stops working, we throw it away and buy a new bottle." With larger compound sets, QC is a much bigger issue. As each compound will most likely be used only once in a given assay, changes in activity over time are impossible to measure. A false negative due to loss of compound in the screening well from precipitation or decomposition will not be detected. Compound redundancy between sets and pharmacological redundancy within a set is a useful safety net but will not help for unique compounds with unique mechanism of action. A general QC method would be very useful to check compound concentrations and identity in screening sets. This is often omitted in the screening process due to issues with compound consumption. Chemogenomics sets are purchased in low individual compound amounts (< 1 ml of 10 mM solutions). This quantity is sufficient for many assays, but will be rapidly depleted if used for QC by NMR or standard LCMS. Considering the effort in assay development, screening and data analysis necessary for target discovery with chemogenomic compound sets, a low consumption, high-sensitivity QC system would be a valuable addition to the University's capabilities.

4. Lead Discovery and Phenotypic Screening

Very few groups in the University are involved directly in lead discovery or phenotypic screening. The necessity to have a large (~100k) diverse compound collection and the infrastructure to maintain and screen the set precludes widespread use. Large diversity sets within the university are summarized in Table 2. Lead discovery is usually performed *via* collaboration with large national screening centres like the Broad Institute, NIH Chemogenomics Centreor other groups with HTS expertise. The newly launched IMI European Lead Factory in which the University is a participant will provide another service centre where groups with a developed assay will be able to do lead discovery for their chemical probe or drug discovery projects.

Name	Number of compounds	Source	Owner	Description
DTP Mechanistic Set	879	DTP	TDI	Representatives from NCI60 compounds with different growth inhibition patterns
DTP Diversity Sets I and II	1900	DTP	TDI	Chemically diverse collection
ChemBridge's DIVERSet	50000	ChemBridge	TDI	Chemically diverse collection
Russell Group Collection	10000	Custom	Russell Group (CRL)	Chemically diverse collection

Table 2. Lead Discovery Compound Sets

The storage of large compound collections is done in a plate-based system and handling is generally done manually. Issues with QC of these collections are similar to chemogenomics sets as described in section *3.2* due to precipitation and decomposition. The problem of library analysis is magnified due to the larger size of the collections. When screening diversity sets for new leads, false negatives are not disastrous as long as at least one new genuine lead is discovered. False positives occur when a decomposition product is active in the assay and can be problematic. Genuine hits need to be confirmed, usually by re-synthesis and confirmation of putative hits.

In terms of assay technology, the University is well equipped in terms of screening equipment. The recent audit done by research services into the University's scientific equipment shows good capabilities in terms of biochemical and biophysical assays (Table 3, compiled from https://www.research-facilities.ox.ac.uk).

Purpose	Type	Number	Description
Sample Handling	Pipetting Plate Makers	25	Prepare 96- and 384-well plates for screening.
Detection	Fluorescence and Multi-Mode Microplate Readers	40	Standard optical assay plate readers.
	Biolayer Interferometers and SPR Spectrometers	7	Protein-ligand interaction measurements.
	NMR Spectrometers (≥500 MHz)	9	Protein-ligand interaction and enzyme inhibition measurements.

Table 3. Screening Equipment Summary

5. Lead Optimisation

For groups involved in drug or probe discovery, lead discovery is followed by further chemical optimisation to deliver a chemical probe or clinical candidate. Optimization will aim to improve potency but also ADME (absorption, distribution, metabolism and elimination) properties and anti-target selectivity to ensure a molecule's utility in cellular or *in vivo* models. There are many ADME and safety *in vitro* assays used in drug discovery and a few of the most common are shown in Table 4. No groups surveyed in the University setup and use drug optimisation assays in their own labs. These assays are general and are routinely run in external commercial labs. There are many companies in the UK who supply these services but Cyprotex is the company most frequently used by University research groups.

Assay Name	Description
HLM	Human liver microsomes are used to predict hepatic metabolism.
CACO-2	Compound flux across a CACO-2 cell monolayer is used to predict intestinal absorption.
MDR1-MDCK	Limited compound efflux measured across a p-glycoprotein over-expressing MDCK cell monolayer precludes CNS penetration and can limit intestinal absorption.
logD	LogD is a measure of a compound's polarity and is a major factor in a compound's potency, selectivity, solubility, permeability, stability and toxicity.

Table 4. Assays Used in Lead Optimization

When a compound has been optimised *in vitro*, it is often used in an *in vivo* model. In order to interpret *in vivo* effects with respect to *in vitro* pharmacology, measuring the pharmacokinetics of the compound is necessary. To progress further towards clinical candidacy it is also necessary to perform animal toxicology studies. A few research groups in the University that we surveyed do *in vivo* pharmacokinetic and toxicology studies either through their own Home Office licences or outsourcing. For groups that do not have these capabilities, Mike Stratford in the Gray Institute has experience and some extra capacity to run pharmacokinetic studies for collaborative projects. There are also many companies that offer these services including Cyprotex in the UK.

6. Data Capture and Sharing

The compounds available and data generated in target discovery and lead discovery screening is a valuable University resource both directly and indirectly. Currently the test data that is generated forms the backbone of biological hypotheses and is shared *via* publications. There are many ways individual research groups capture such data within the University but the data is not shared, often for technical reasons. This data is very valuable both in the knowledge derived from it but also in terms of sharing expertise and capabilities. It is currently very difficult for one research group within the University to know if another group has expertise in a certain assay or screening technique. There is also no mechanism for one group to know if another group has compounds of interest for either target discovery or lead discovery.

7. Metabolomics

Metabolomics is an umbrella which covers many potential studies. The researchers interviewed for this audit have interests in scenarios which vary in terms of sources of metabolites, metabolites measured, and detection method (Table 5). There is general agreement among most of the research groups and departments approached that capability in metabolomic analysis is a major gap in the University's research infrastructure. These types of studies are not easily outsourced to commercial suppliers as the cost of the analysis is prohibitive and interpretation of the results requires dedicated expertise.

Method	Biophase		Metabolites													
			Amino Acids		Carbo-hydrates		Lipids		Nucleo-tides		TCA		Cofactors		Other	
NMR	*in vitro*	extra-cellular														
MS	*in vitro*	cell extract	A	B C	A	B C	A	B C	A	B C D	A	B C	A	B C	A	B C
	in vivo	urine	E		E		E		E		E		E		E	
	in vivo	blood	E		E		E		E		E		E		E	
	in vivo	tissue														

Table 5. Metabolomic analysis. Research Groups: **A** NMR facility in Chemistry Department, **B** MS facility in Chemistry Department, **C** MS group in Centre for Cellular and Molecular Physiology, **D** Gray Institute, **E** Clinical Trial Service Unit.

There are at least five groups doing metabolomics to some extent within the University (Table 5, **A**-**E**). The CTSU is using NMR to measure metabolites in blood and urine derived from patients participating in clinical trials. In the Chemistry Department, bespoke collaborations have formed around the NMR and MS groups. In the CCMP, MS is used to measure metabolites in collaborative projects. In the Gray Institute, nucleotides are measured in cell extracts. None of these groups have the capacity in terms of equipment or expertise to extend their capabilities much beyond current projects and would not be able to satisfy the rest of the University's demand without expansion.

8. Strategic Recommendations

8.1 Screening Set QC

There is a need for a dedicated and automated low consumption, high-sensitivity LCMS system for QC of compounds in screening sets. Although a large number of LCMS systems are present in the University, there are none dedicated to the QC of the University's larger compounds collections as described in sections *3.1* and *4*. A small molecule dedicated system would minimise time loss in developing new methods. The system should be a nano-LC system and should have the highest possible sensitivity in UV/vis and MS detection modules. An additional ELSD module would also allow estimation of compound concentration. The autosampler needs to be compatible with 96- and 384-well screening plates and have stacking capabilities to allow unattended analysis of multiple plates. The expertise to house and run the QC system exists in the TDI and Chemistry Department.

8.2 Compound and Screening Database

There is a large communication gap in the University around how different groups use small molecules. Hopefully, this audit will serve as a step in bridging the gap, but there is need of a longer term solution. It would be beneficial for many researchers to know what compounds are available in the University, both in terms of chemical structure and pharmacological profile. Many researchers approached for this audit would like to see a University-wide database of compounds and assay data. Such a database would serve many purposes including the ability to find what compounds and associated pharmacological profile are available for testing, what groups have expertise in certain assays, and which groups are working on similar biological pathways where synergy may be possible. With the recent HEFCE funding for the Big Data Institute in the Medical Sciences Division, a database of compounds and pharmacology will complement clinical, metabolomics and genomic data capabilities.

This need has been addressed recently with Wellcome Trust ISSF and John Fell Fund awards to establish a compound and screening database.

8.3 Outsourcing Lead Optimisation Assays

There are many commercial suppliers of *in vitro* lead optimisation assays and *in vivo* studies as outlined in section *5*. Research groups that need access to these capabilities may wish to contact Cyprotex as they have developed a good working relationship with the University. There are also a number of reliable suppliers in India (GVK Biosciences) and China (Wuxi AppTec). For a more complete drug discovery service, the UK has a number of commercial suppliers such as BioFocus, Evotec and Argenta.

8.4 Metabolomics Capabilities

As outlined in section *7*, there is great interest from the University's research groups in increasing metabolomics capabilities. Metabolomics covers many method and molecule combinations (Table 5), and there is not a simple solution to address these needs. There is also general consensus that there is a gap in terms of analytical equipment and expertise in equal measure. Any increases in capital equipment for analytical analysis would need to be matched with technical expertise or the benefit would be minimal. The large amount of data generated in a metabolomic analysis needs to be interpreted with an understanding of both the analytic methods and underlying biology. Groups that benefit the most from metabolomics research operate in a collaborative manner with

metabolomics specialists. For metabolomics analysis of clinical samples there also needs to be an investment in long term storage of samples.

The need for greater metabolomics capabilities will continue to increase now that the Medical Sciences Division has been awarded HEFCE funding for the Big Data Institute. The pairing of genomic and metabolomic information will be critical in getting maximum benefit from the data collected from clinical trials across the division. As recommended in the recent ISSF mass spectrometry audit from Benedikt Kessler, to expand metabolomics capabilities in the University it would be necessary to recruit at least one faculty level researcher to establish a metabolomics centre in the University. Julian Griffin from Cambridge was suggested by several of the researchers surveyed for this audit as someone with the necessary expertise and experience in data generation and analysis.

9. Appendix

Table A1. Departments contacted for this audit.

Division	Sub-division	Department
MPLS		Chemistry
		Computer Science
		Maths
		Physics
Medical Sciences	Non-Clinical	Dunn School of Pathology
	Non-Clinical	DPAG
	Non-Clinical	Pharmacology
	Non-Clinical	Biochemistry
	NDM	SGC
	NDM	Ludwig Foundation
	NDM	Stone Group
	NDM	CCMP
	NDM	CTSU
	NDM	Target Discovery Institute
	NDM	WT Centre for Human Genetics
	NDM	NDM Experimental Medicine
	NDM	STRUBI
	Oncology	Molecular Oncology
	NDORMS	Botnar Research Institute
	NDORMS	Kennedy Institute
	NDCN	Clinical Neurology
	RDM	Cardiovascular Medicine
	RDM	Nuffield Division of Clinical Laboratory Sciences
	RDM	Investigative Medicine Division
	RDM	OCDEM

Black – responses were used in this survey. Blue – all research activities out of the scope of this audit. Green – no departmental representative was available for survey.

Abbreviations: MPLS – Division of Mathematics, Physics and Life Sciences; NDM – Nuffield Department of Medicine; NDORMS – Nuffield Department of Orthopaedic and Rheumatological

Medical Sciences; NDCN – Nuffield Department of Clinical Neurology; RDM – Radcliffe Department of Medicine; DPAG – Department of Physiology Anatomy and Genetics; SGC – Structural Genomics Consortium; CCMP – Centre for Cellular and Molecular Physiology; CTSU – Clinical Trial Service Unit; STRUBI – Structural Biology; OCDEM – Oxford Centre for Diabetes, Endocrinology and Metabolism.

Figure A1. Overlap in representative chemogenomics screening sets. **A.** Percent overlap between all sets. **B.** Total and unique compounds tested with increasing number of sets tested.

A

fraction of overlap	BML-2840	BML-2842	LOPAC	Prestwick	Spectrum
Spectrum	1%	6%	3%	22%	100%
Prestwick	1%	19%	3%	100%	16%
LOPAC	14%	3%	100%	3%	2%
BML-2842	2%	100%	1%	10%	2%
BML-2840	100%	2%	5%	0%	0%

B